iB2004|050035

RECEIVED

1 0 FEB 2004

WIPO   PCT

Bescheinigung   Certificate   Attestation

Die angehefteten Unterla-
gen stimmen mit der
ursprünglich eingereichten
Fassung der auf dem näch-
sten Blatt bezeichneten
europäischen Patentanmel-
dung überein.

The attached documents
are exact copies of the
European patent application
described on the following
page, as originally filed.

Les documents fixés à
cette attestation sont
conformes à la version
initialement déposée de
la demande de brevet
européen spécifiée à la
page suivante.

Patentanmeldung Nr.   Patent application No.   Demande de brevet n°

03075199.4

PRIORITY
DOCUMENT
SUBMITTED OR TRANSMITTED IN
COMPLIANCE WITH RULE 17.1(a) OR (b)

Der Präsident des Europäischen Patentamts;
Im Auftrag

For the President of the European Patent Office

Le Président de l'Office européen des brevets
p.o.

R C van Dijk

Anmeldung Nr: Anmeldetag:
Application no.:   03075199.4   Date of filing: 20.01.03
Demande no: Date de dépôt:

Anmelder/Applicant(s)/Demandeur(s):

Koninklijke Philips Electronics N.V.
Groenewoudseweg 1
5621 BA   Eindhoven
PAYS-BAS

Bezeichnung der Erfindung/Title of the invention/Titre de l'invention:
(Falls die Bezeichnung der Erfindung nicht angegeben ist, siehe Beschreibung.
If no title is shown please refer to the description.
Si aucun titre n'est indiqué se referer à la description.)

Video coding

In Anspruch genommene Priorität(en) / Priority(ies) claimed /Priorité(s)
revendiquée(s)
Staat/Tag/Aktenzeichen/State/Date/File no./Pays/Date/Numéro de dépôt:

Internationale Patentklassifikation/International Patent Classification/
Classification internationale des brevets:

H04N7/26

Am Anmeldetag benannte Vertragstaaten/Contracting states designated at date of
filing/Etats contractants désignées lors du dépôt:

  AT BE BG CH CY CZ DE DK EE ES FI FR GB GR HU IE IT LU MC NL
  PT SE SI SK TR LI

Video Coding

The invention relates to video coding

1. Background art

1.1 Relevance and applications of the upcoming H.264 standard

During the recent years, a new ITU-T specification for video coding has been developed – H.26L, which has become broadly recognized for offering superior coding efficiency in comparison with the existing standards ("same signal-to-noise ratio for up to 50% less bits"). Although the gain of H.26L generally decreases in proportion to the picture size, the potential for its deployment in a broad range of applications is undoubted. This potential has been recognized through formation of the so-called Joint Video Team (JVT), which has the task to finalize H.26L as a new joint ITU-T/MPEG industrial standard. The new standard is expected to be formally approved in 2003 as ITU-T H.264 or ISO/IEC MPEG-4 AVC (Advance Video Coding). In the meantime, H.264-based solutions are being considered in other standardization bodies, such as the DVB, DVD Forum and Blu-ray disk consortium, while SW/HW implementations of H.264 encoder/decoder are already becoming available. The development of H.264 is reflected in publicly accessible JVT documents [1], including the complete draft of the standard.

1.2 Particularities of H.264 syntax and coding tools

H.264 employs same principles of block-based motion-compensated hybrid transform coding that are known from the established standards such as MPEG-2. The H.264 syntax is, therefore, organized as the usual hierarchy of headers such as picture-, slice- and macro-block headers, and data such as motion-vectors, block-transform coefficients, quantizer scale, etc. Nevertheless, new syntax and coding methods are introduced at both the header- and the data level. A brief summary of some main particularities of H.264 is given below. The most relevant (for our proposal) particularities are subsequently explained in more detail in separate sections, taking [1] as reference. Typical block-diagrams illustrating H.264 encoding and decoding are given in Appendix A as Figures 1a and 1b.

1.2.1   A brief overview

H.264 separates the Video Coding Layer (VCL), which is defined to efficiently represent the content of the video data, and the Network Adaptation Layer (NAL), which formats data and provides header information in a manner appropriate for conveyance by the high level system. One of the main particularities of H.264 at the data level is the use of more elaborate partitioning and manipulation of 16x16 macro-blocks. In H.264, the motion compensation process can form segmentations of a macro-block as small as 4x4 in size, using motion vector accuracy of one-fourth or one-eight of a sample grid. Also, the selection process for motion compensated prediction of a sample block can involve a number of stored previously decoded pictures, instead of only the adjoining ones. Even with intra coding, it is now possible to form a prediction of a block using previously decoded samples, in that case from the same picture. The rules for this spatial-based prediction are described by the so-called intra prediction modes. After motion compensated- or spatial-based prediction, the resulting prediction error is normally transformed and quantized based on 4x4 block size, instead of the traditional 8x8 size. An additional provision called Adaptive Block Transform has been

considered, which allows to use multiple transforms to match the possible sizes of prediction blocks. But it is not yet clear whether this tool will be included in the final H.264 specification. The H.264 also uses new concepts in other coding stages. For example, H.264 departs from the usage of the DCT (Discrete Cosine Transform), which is used in previous standards such as MPEG-2. It also specifies different rules and designs for operations such as Entropy Coding or VLC (Variable Length Coding), quantization, etc. But, in contrast to the earlier explained concepts, most of these concepts only allow fixed implementation and are described by syntax elements which cannot be set-up below the sequence-, GOP- or picture level.

1.2.2   Motion compensation

Most established video coding standards (e.g. MPEG-2) inherently use block-based motion compensation as a practical method of exploiting correlation between subsequent pictures in video. This method attempts to predict each macro-block in a certain picture by its "best match" in an adjacent reference picture. This prediction is usually performed using only 16x16 luminance blocks, and the results of it are then also applied to the corresponding chrominance pixels. If the pixel-wise difference between a macro-block and its prediction is small enough, the prediction error, i.e. the difference between a macro-block and its prediction is encoded rather that the macro-block itself. The relative displacement of the prediction block with respect to the coordinates of the actual macro-block is indicated by a motion vector, which is coded separately. Figure 2 illustrates the case of bi-directional prediction, where two reference pictures are used, one in the past and one in the future. Pictures that are predicted in this way are called B-pictures. Otherwise, pictures that are predicted only from past pictures are called P-pictures.
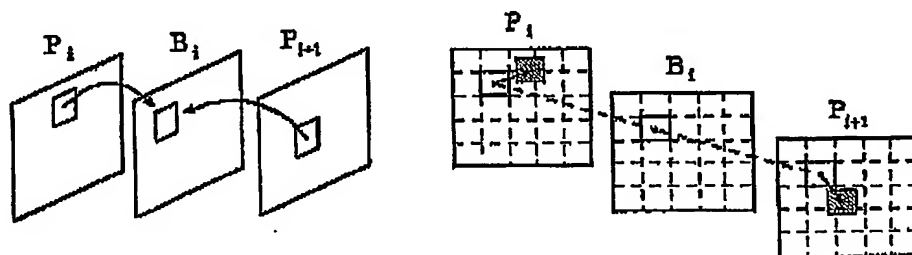

Figure 2. Each macro-block in a B-frame can be predicted from a block from the past P-frame, or one from the future P-frame, or by an average of two blocks, each from a different P-frame.

Much of the bit-rate savings offered by H.264 can be actually attributed to its improved methods of motion compensation. This is explained in more detail in the following subsection.

1.2.2.1 Multiple prediction block sizes in H.264

In H.264, variable block size can be used for inter-, i.e. temporal prediction of a macro-block. Accordingly, a macro-block can be partitioned into a number of smaller blocks and each of these sub-blocks can be predicted separately (the prediction is still performed using only luma 16x16 blocks). Hence, different sub-blocks can have different motion vectors and can even be retrieved from different reference pictures (see Section 1.2.2.2). The number, size and orientation of prediction blocks is uniquely determined by definition of *inter prediction modes*, which describe possible partitioning of a macro-block into 8x8 blocks and

further partitioning of each its 8x8 sub-block. This is also shown in Figure 3. The H.264 syntax includes elements such as *mb_type* and *sub_mb_type* to indicate to a decoder which partition has been used with a certain macro block for the inter prediction. This is explained in more detail in Section 7.4.5 (Tables 7-13, 7-12, 7-16, 7-17) in [1].
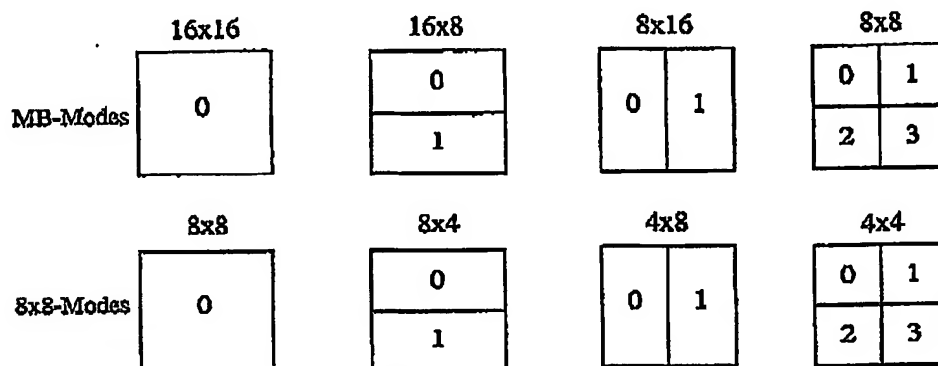

Figure 3. H.264 allows inter prediction of a macro-block based on separate partition of its sub-blocks.

1.2.2.2 Multiple reference pictures

In H.264, inter prediction for a certain macro-block can be formed by also taking blocks from more distant previously decoded future- or past pictures, instead only from the adjoining ones. This is referred to as *multiple reference pictures* and is illustrated in Figure 3. The selection of a certain reference picture for prediction of a sub-block in a macro-block (see previous section) is indicated in the bitstream by the value of syntax elements *ref_idx_l0* and *ref_idx_l1* ([1], Sec. 7.4.5.1)
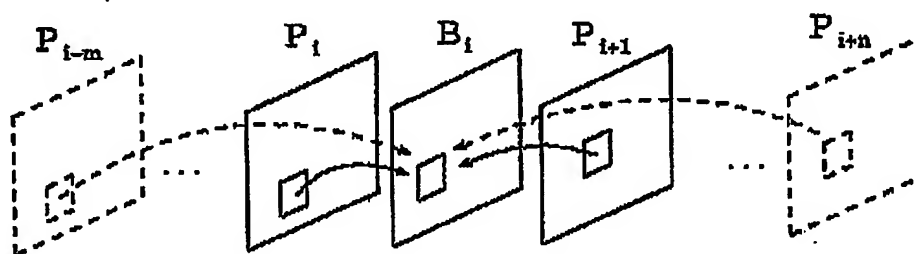

Figure 4. Illustration of the multiple reference pictures prediction, for the case of bi-directional prediction.

1.2.3 De-blocking filter

In H.264 conditional filtering is applied to all macro-blocks of a picture. For luma, as the first step, the 16 samples of the 4 vertical edges of the 4x4 raster shall be filtered beginning with the left edge, as shown in Figure 5. Filtering of the 4 horizontal edges (vertical filtering) follows in the same manner, beginning with the top edge. The same ordering applies for chroma filtering, with the exception that 2 edges of 8 samples each are filtered in each direction. For each boundary between neighbouring 4x4 luma blocks, a "Boundary Strength" Bs is assigned. If Bs=0, filtering is skipped for that particular edge. In all other cases filtering is dependent on the local sample properties and the value of Bs for this particular boundary segment ([1], Sec. 8.7). Several syntax elements are used to indicate in the bitstream whether the deblocking filter shall be applied to the edges controlled by the macro-blocks within the

current   slice   and   with   which   parameters.   Such   elements   are   e.g.
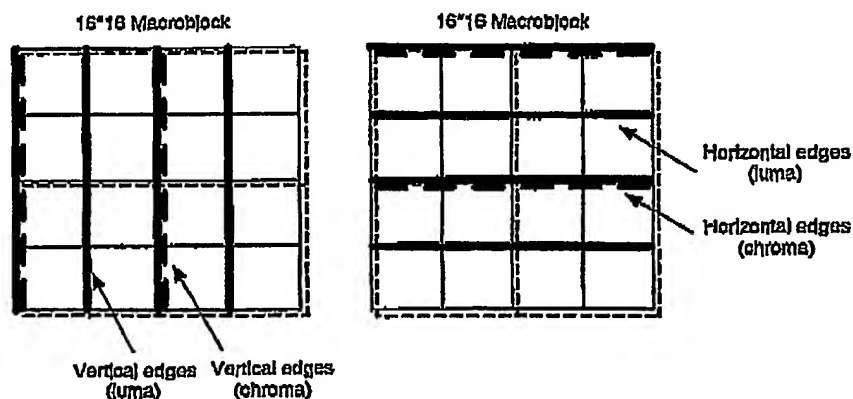*disable_deblocking_filter_flag* and *slice_alpha_c0_offset_div2* ([1], Sec. 7.4.3)

Figure 4. The de-blocking filtering is applied along several boundaries of a macro-block and within its sub-blocks

1.2.4   Adaptive Block Transform

In H.264 the residual coding is by default performed using a 4x4 integer transform, which is similar but not compatible with the DCT (Discrete Cosine Transform) used in MPEG-2. Hence, the prediction error (i.e. the pixel-wise difference between a macro-block and its prediction) is divided into 16 luma 4x4 blocks and 8 chroma 4x4 blocks, as shown in Figure 5. After the transformation, one DC coefficient is obtained for each 4x4 block, which leaves 16 DC coefficients for the luma and 4 DC coefficients for each component of the chroma. The chroma DC coefficients are then grouped and transformed again, using another 2x2 transform.

Luma residual coding 4x4 block order   Chroma residual coding 4x4 block order

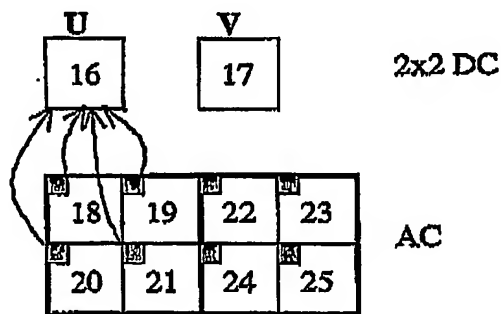| 0	1	4	5
2	3	6	7
8	9	12	13
10	11	14	15

Figure 5 Illustration of 4x4 residual coding order in H.264

In recent drafts of H.264 transforms of size 4x8, 8x4, and 8x8 have been specified, in addition to the default 4x4 transform. This feature is called Adaptive Block Transform (ABT) and applies to the luma residual (the chroma residual coding process therefore

remains as described above). The use of ABT is indicated in the bitstream by a a parameter called *adaptive_block_size_transform_flag* ([1], Section 12). In the case of inter coding, the size of a particular transform size will coincide with the block size used for prediction (see Section 1.2.2.1). For intra macroblocks, the block size used for intra prediction is connected to the block size of the transformation. The order of the assignments of syntax elements for luma resulting from coding a macroblock to sub-blocks of the macroblock if the ABT features is used is shown in Figure 6. A 8x8 block may contain 1, 2, or 4 transform blocks. An indication that an 8x8 block contains coefficients means that the 8x8 transform blocks or one or more of the 2, or 4 transform blocks within the 8x8 block contains coefficients. More details about the syntax and semantics of ABT can be found in Section 12 in [1].

CBPY 8x8 block order

Luma residual coding ABT block order
for one CBPY 8x8 block

Figure 6 Ordering of blocks of CBPY (Coded Block Pattern) and luma residual coding of ABT blocks

2. Problem statement

One of the main purposes of development of H.264 was to respond to the growing need for substantially higher compression of moving pictures for applications such as video conferencing, internet streaming and communication, etc. Therefore, H.264 includes several coding tools that are suited for smaller picture formats and low bitrates being characteristic for such applications, but become less effective with the increase of the picture size. This is also confirmed by experiments with High Definition (HD) video, where it is generally observed that, at a certain point, an increase of the bitrate does not give a proportional increase of the picture quality in the situation where all the characteristic H.264 coding tools are enabled. In other words, even though some H.264 coding tools are responsible for achieving good picture quality at remarkably low bitrates, they seem less contributing, of even disturbing at higher bitrates. This implies that the typical H.264 operation can be inadequate for applications where bit rate constraints need not be as tight, yet virtually transparent picture quality should be achievable (As in the case of de-blocking filtering, the H.264 syntax allows conditional operation of certain coding tools. However, in practical automated encoding, these conditions are determined by local low-level computations that usually attempt to minimize the bitrate rather than to preserve the picture quality).. Such an application is distribution of HD movies on discs with high storage capacity such as Blu-ray Disk (25GB, 0.1 mm cover layer) or Blue DVD (15GB, 0.6 mm cover layer).

A particularly relevant problem of H.264 in this application area is that it has the tendency to remove the film grain, which effect is hardly reduced even when the bitrate is considerably increased, in the situation where typical H.264 coding settings used. The film grain refers to (slightly visible) noise that is introduced in film due to imperfection of recording equipment

and environment, but has become so common that it is generally expected and is often even preferred by directors as a means for achieving a natural "film look".

3. Summary of the invention

An object of the invention is to provide better quality for higher bit rates of a given coding standard. To this end, the invention provides a method of coding, an encoder, a coded bit-stream and a record carrier as defined in the independent claims. Advantageous embodiments are defined in the dependent claims.

According to a first aspect of the invention, the coding disables some of the tools provided by the given coding standard, wherein an identification of the disabled tools is included in the bit-stream, the disabled tools being one or more out of the group of:
— bidirectional predictive coding of pictures or picture parts
— use of a de-blocking filter
— use of more than one reference picture.

By providing an identification of the disabled tools, the encoder signals to a decoder that the disabled tools are not used. In the case the coding standard provides parameters or indicators that can be used to indicate disabled tools, the coded bit-stream can be implemented such that it remains compatible with the standard. In an embodiment, Adaptive Block transforms are used.

Embodiments of the invention are described in relation to the H.264 standard although the invention is also applicable to other coding standards.

According to an embodiment of the invention, a HQ-HD profile of H.264 is proposed that can be used for high quality (virtually transparent) HD video compression, as intended for applications such as publishing of HD movies on high capacity digital carriers such as "Blu-ray disk". Out of the many tools possible and allowed by the H264 standard, only a very specific combination makes it possible to achieve at relative high bit-rates virtually transparent HDTV picture quality. This profile is obtained by selective exclusion of several standard H.264 coding tools or modes that we have found to be not contributing or disturbing for preserving virtually transparent picture quality at higher bit-rates. This exclusion can be easily indicated in the H.264 bit-stream, by enforcing or constraining certain values for several H.264 syntax elements. The benefit of such constraint of H.264 would not only be in that it would create unique conditions for approaching transparent picture quality while using H.264, but also in that it would also enable construction of less complex H.264 encoders and decoders for this purpose.

In this embodiment, the following mandatory exclusions/constraints of the standard coding tools that would uniquely define a profile:
— Exclusion of B pictures / B slices (Section 10 in [1])
— Exclusion of the de-blocking filter (Section 1.2.3)
— Exclusion of at least one of the block sizes for inter prediction which are smaller than 8x8 (Section 1.2.2.1)
— Constraining the number of reference pictures to be used for prediction to 1 (Sec. 1.2.2.2)
— Although currently ABT is most likely not to be part of the 1st version of the standard, we propose to include ABT to the HQ-HD profile of H.264 (*the use of ABT (see section 1.2.4) that has existed in H.264 so far, but is currently being considered for exclusion from the final specification)*

In addition to the disabling of standard H.264 coding tools and modes, the proposal also includes recommendation related to the use of some non-normative parts of H.264. One of them is the use of rate-distortion optimization, which is implemented in the JVT test software of H.264 encoder. In this respect, we recommend to not implement any kind of rate-distortion optimization in the H.264 encoder

Embodiments of the invention can directly be implemented in a standard encoder such as the H.264 encoder shown in Fig. 1a. Further, because it is not necessary for the encoder to be capable of using the disabled tools (e.g. for another operation mode), it is possible to provide a simple encoder with a reduced set of tools in combination with some means to include the correct parameters in the bit-stream to identify the disabled tools. As far as the disabled tools concern tools for which the standard provides an indicator indicating that the tool is not used, the simple encoder provides a compatible bit-stream.

4. Practical embodiment

The following selective use of the tools of H.264 can provide almost transparent quality at bitrates of ~15Mbs:

H.264 tools	reference all tools enabled	preferred embodiment
GOP length	12-24	12
Number of B frames	1 or 2	0
Q_par (B)	Q_par(P)+1	not applicable
De-blocking filter	Enabled	Disabled
RD optimization	Enabled	Disabled
InterSearch 16x16	Enabled	Enabled
InterSearch 16x8	Enabled	Enabled
InterSearch 8x16	Enabled	Enabled
InterSearch 8x4	Enabled	Disabled
InterSearch 4x8	Enabled	Disabled
InterSearch 4x4	Enabled	Disabled
Number Reference Frames	2-5	1

The use of Adaptive Block Transforms is preferred.

Appendix B shows a comparison of the reference with the preferred embodiment indicating that the preferred embodiment leads to a significant increase in quality.

It should be noted that the above-mentioned embodiments illustrate rather than limit the invention, and that those skilled in the art will be able to design many alternative embodiments without departing from the scope of the appended claims. In the claims, any reference signs placed between parentheses shall not be construed as limiting the claim. The word 'comprising' does not disable the presence of other elements or steps than those listed in a claim. The invention can be implemented by means of hardware comprising several distinct elements, and by means of a suitably programmed computer. In a device claim enumerating several means, several of these means can be embodied by one and the same item of hardware. The mere fact that certain measures are recited in mutually different

PHNL030092EP-Q 8

dependent claims does not indicate that a combination of these measures cannot be used to advantage.

5. References

[1] "Joint Final Committee Draft (JFCD) of Joint Video Specification (ITU-T Rec. H.264 | ISO/IEC 14496-10 AVC)" – file "JVT-D157.doc" at ftp://ftp.imtc-files.org/jvt-experts/

[2] H.264 tutorials at: http://www.vcodex.fsnet.co.uk/h264.html

CLAIMS:

1.   A method of coding a video signal according to a predefined standard, wherein in a
given operation mode some of the tools provided by the predefined standard are disabled,
and wherein an identification of the disabled tools is included in the bit-stream, the disabled
tools being one or more out of the group of:
— bidirectional predictive coding of pictures or picture parts
— use of a de-blocking filter
— use of more than one reference picture.

2.   An encoder comprising
        means for coding a video signal according to a predefined standard, wherein in a
given operation mode some of the tools provided by the predefined standard are disabled,
        means for including an identification of the disabled tools in the bit-stream,
the disabled tools being one or more out of the group of:
— bidirectional predictive coding of pictures or picture parts
— use of a de-blocking filter
— use of more than one reference picture.

3.   A coded bit-stream representing a video signal, the bit-stream including an
identification of disabled tools, which disabled tools were disabled in the coding of the coded
bit-stream, the disabled tools being one or more out of the group of:
— bidirectional predictive coding of pictures or picture parts
— use of a de-blocking filter
— use of more than one reference picture.

4.   A record carrier having stored thereon a coded bit-stream as claimed in claim 3.
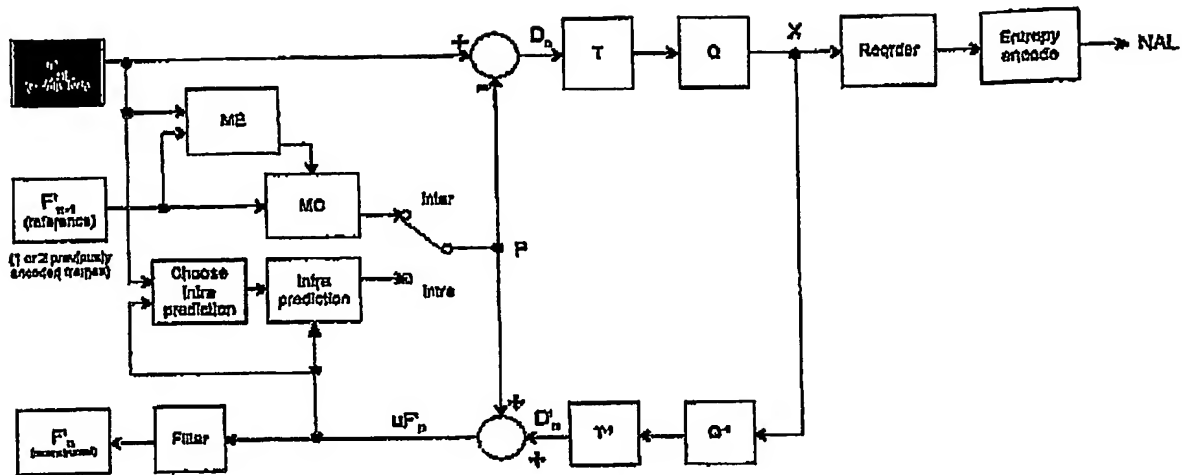
PHNL030092EP-Q   10

Appendix A

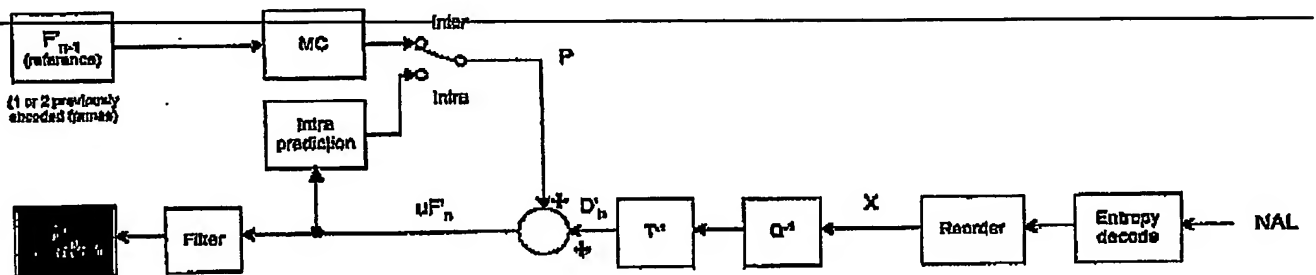Figure 1a. Block diagram of H.264 encoder (taken from [2])

Figure 1b. Block diagram of H.264 decoder (taken from [2])

- ME = Motion Estimation
- MC = Motion Compensation
- Q = Quantization
- Q^{-1} = Inverse Quantization
- T = Transform
- T^{-1} = Transform,
- Filter = de-blocking filter
- F_{i-1} = i-th reference picture for inter prediction
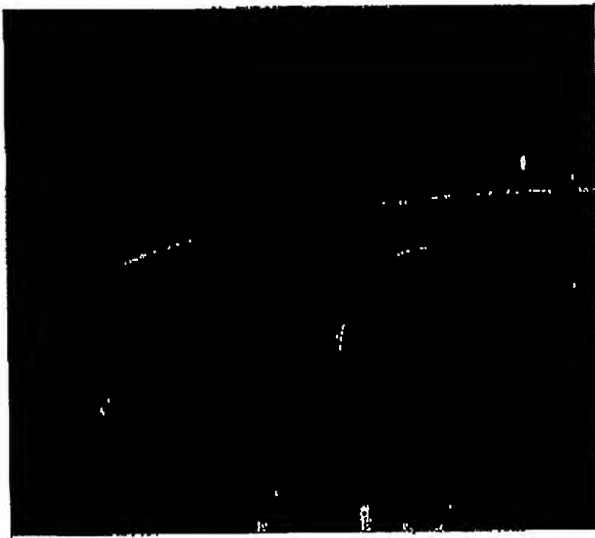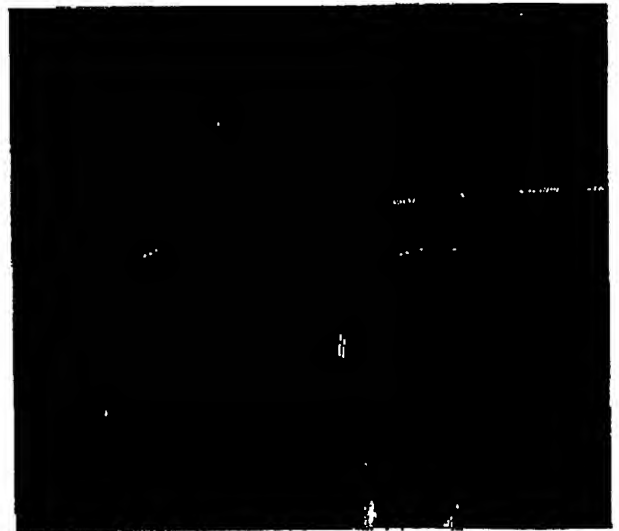- NAL = Network Abstraction Layer (see Section 1.2.1)

PHNL030092EP-Q 11

Appendix B

Original

Reference at 13 Mbs

Preferred embodiment at 14 Mbs

PCT Application

PCT/IB2004/050035